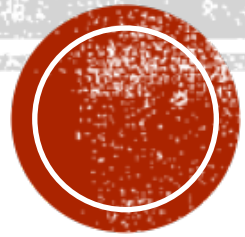
LOAD BALANCING, AUTOSCALING, SERVERLESS COMPUTING

-Anilkumar Dantala
-Ankit Patel
-Ketul Shukla



LOAD BALANCING

- Azure – Load Balancer
  - Azure Load Balancer supports TCP/UDP-based protocols such as HTTP, HTTPS, and SMTP, and protocols used for real-time voice and video messaging applications.
  - High availability and robust performance for your applications
  - Internal load balancer for traffic between virtual machines inside your private virtual networks
  - Load Balancer probes the health of your application instances, automatically takes unhealthy instances out of rotation, and reinstates them when they become healthy again

- Google – Cloud Load Balancinng
  - Cloud Load Balancing can put your resources behind a single anycast IP and **scale your resources up or down with intelligent Autoscaling.**
  - You can apply Cloud Load Balancing to all of your traffic: HTTP(S), TCP/SSL, and UDP. You can also terminate your SSL traffic with HTTPS Load Balancing and SSL proxy.
  - It supports 1 Million+ queries per second with consistent high performance and low latency.
  - Internal Load Balancing enables you to build scalable and highly available internal services for your internal client instances without requiring your load balancers to be exposed to the Internet.

- AWS – Elastic Load Balancing
  - Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones.
    - Application Load Balancer - routes traffic to targets within VPC based on the content of the request, operates at individual request layer (Layer 7) , HTTP and HTTPS traffic
    - Network Load Balancer - routes traffic to targets within VPC and is capable of handling millions of requests per second while maintaining ultra-low latencies, Operates at layer 4 ,TCP traffic
    - Classic Load Balancer - provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level
  - High Available, Secure, Elastic, Flexible, Robust monitoring and auditing

AUTO SCALING

- Azure
- Azure provide a common set of autoscaling functionality for VM Scale Sets, Azure App Service, and Azure Cloud Service. Scaling can be performed on a schedule, or based on a runtime metric, such as CPU or memory usage.
- You can configure autoscaling by using PowerShell, the Azure CLI, an Azure Resource Manager template, or the Azure portal
- Basically, there are two main ways to scale resources:

  Vertical: Scaling up and down

  Horizontal: Scaling out and in

Google

- Compute Engine offers auto scaling to automatically add or remove virtual machines from an instance group based on increases or decreases in load.
- You just define the auto scaling policy and the autoscaler performs automatic scaling based on the measured load.
- Policies –
  - CPU utilization – It is the most basic autoscaling that you can perform.
  - Load balancing serving capacity - Set up an autoscaler to scale based on load balancing serving capacity and the autoscaler will watch the serving capacity of an instance group.
  - Stackdriver Monitoring metrics - If you export or use Stackdriver Monitoring metrics, you can set up autoscaling to collect data of a specific metric and perform scaling based on your desired utilization level

AWS

- AWS Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost.
- Benefits
  - SETUP SCALING QUICKLY
  - MAKE SMART SCALING DECISIONS
  - AUTOMATICALLY MAINTAIN PERFORMANCE
  - PAY ONLY FOR WHAT YOU NEED

SERVERLESS COMPUTING

Serverless computing, or FaaS (Functions-as-a-Service) lets developers focus on building event-based applications on a function by function basis while it takes care of deploying, running and scaling the code

AZURE FUNCTION

- Microsoft allows you to create functions in native languages, such as C# and Javascript
- Azure has easy integrations with a range of external services, such as VS Team Services, Bitbucket and Github, allowing the deploying of code in the cloud
- Microsofts Logic apps, allows you to set and manage data processing tasks and to assign workflow paths.
- Azure's cloud function usage is billed in the same way as Amazon, calculating the total cost from both number of triggers and execution time.

GOOGLE CLOUD FUNCTION

- Google was the last of the big three to add a serverless option
- functions only to be written in JavaScript and executed in a standard Node.js runtime environment
- Google currently restricts projects to having fewer than 20 triggers
- After one million free requests, its prices are double those of AWS and Microsoft: $0.04 for 100,000 invocations, plus $0.04 for 100,000 milliseconds

AWS LAMBDA

- To activate Lambda, just upload your code and Lambda takes care of everything required to run and scale your code with high availability.
- native support for a range of runtime environments, including JavaScript, Python, NodeJS and C#
- pricing for Lambda, you only pay for the compute time you consume

REFERENCES

- https://azure.microsoft.com/en-us/services/load-balancer/
- https://cloud.google.com/load-balancing/
- https://aws.amazon.com/elasticloadbalancing/
- https://www.bizety.com/2018/01/09/serverless-comparison-aws-lambda-vs-microsoft-azure-functions-vs-google-cloud-functions/